LocalRank: Ranking Web Pages Considering Geographical Locality by Integrating Web and Databases

Jianwei Zhang[1], Yoshiharu Ishikawa[1,2],
Sayumi Kurokawa[1], and Hiroyuki Kitagawa[1,2]

[1] Department of Computer Science, Graduate School of Systems and Information Engineering
[2] Center for Computational Sciences
University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan
{zjw,saku39}@kde.cs.tsukuba.ac.jp,
{ishikawa,kitagawa}@cs.tsukuba.ac.jp

Abstract. In this paper, we propose a method called LocalRank to rank web pages by integrating the web and a user database containing information on a specific geographical area. LocalRank is a rank value for a web page to assess its relevance degree to database entries considering geographical locality and its popularity on a local web space. In our method, we first construct a linked graph structure using entries contained in the database. The nodes of this graph consist of database entries and their related web pages. The edges in the graph are composed of semantic links including geographical links between these nodes, in addition to conventional hyperlinks. Then a link analysis is performed to compute a LocalRank value for each node. LocalRank can represent user's interest since this graph effectively integrates the web and the user database. Our experimental results for a local restaurant database shows that local web pages related to the database entries are highly ranked based on our method.

1 Introduction

Recently, *local search* [1, 2], that provides information on a specific geographical area, has attracted a lot of research interests. For this purpose, gathering web pages related to a specific area becomes an essential task, but so far it is common that local web pages are collected manually. It is also difficult for a computer to determine whether a web page describes information on the area where a user is interested in. The reason is that vagueness often exists in geographical descriptions on web pages. Examples of vague descriptions are as follows: two or more locations may have a same place name, a person name appearing on a web page is mistaken for a place name, and so on.

We also notice another problem that a web page with high popularity on the global web may not be an important one for a specific geographical area. Conversely, an important web page for a specific geographical area may not be ranked to the top place by a conventional web mining method that considers reputation on the global web. Furthermore, for a specific geographical area, there are usually many web resources corresponding to entities existing in the real world (e.g., web pages of a certain organization in this area). They may have been known by a user or can be acquired easily from some web directories or by following the links of a portal web site about this area. For example, consider a user is interested in information about restaurants in Tsukuba city, Japan and he or she has a table shown in Fig. 1. Effective use of this known information to rank web pages is the main concern of this paper.

ID	Name	URL	Address	Phone	Zip Code
1	COCCOLINO	http://coccolino.jp/	3-1-5 Chikuho, Tsukuba, Ibaraki	029-864-4555	300-3257
2	GURUMAN	http://www.omisemall.com/goru/	3-7-17 Azuma, Tsukuba, Ibaraki	029-851-6107	305-0031
3	RANTEI	http://e-tsukuba.jp/rantei/index.htm/	1055-11 Shimohirooka, Tsukuba, Ibaraki	029-851-2603	305-0043
⋮	⋮	⋮	⋮	⋮	⋮

Fig. 1. Restaurants in Tsukuba city

Recently some researches [3, 4] proposed to consider relationships between database entries as a kind of "links" and apply a link analysis to a database. We extend this idea to integrate the web and a user database and perform a link analysis to rank related web pages. In our method, we first construct a linked graph structure for the specific geographical area using entries contained in the database. The nodes of this graph consist of database entries and their related web pages. The edges in the graph are composed of semantic links including geographical connections between these nodes, in addition to conventional hyperlinks. Then a link analysis is performed based on this graph. This graph represents the information of the user provided database, its related web pages, and their relationships. Therefore, the results of the link analysis, called *LocalRank*, not only reflect web pages' popularity on the locally constructed web space (as opposed to the global web), but also take their relevance degrees to the user database into consideration. Hence, our method can be thought of as a ranking approach based on user's interest. Notice that we consider geographical locality by generating geographical links between database entries and between database entries and web pages, instead of directly judging which geographical area a web page intends to describe. In this paper, we also introduce a simple but effective method about how to collect web pages related to a user database.

The remaining part of this paper is organized as follows. Section 2 reviews the related work. The detailed proposed method is presented in Section 3. Section 4 shows the experimental results based on our method. Finally, we conclude this paper and discuss the future work in Section 5.

2 Related Work

Link analysis Currently, link analysis [5, 6], that uses the hyperlinked structure of the web, is an important technology to identify high-quality web pages. Among existing approaches, Kleinberg's *HITS* [7] and Google's *PageRank* [8] are the most representative algorithms. HITS takes a subset of a web graph and generates hub and authority scores for each page in the subset. PageRank enforces that pages are important if important pages link to them. Our LocalRank calculation is similar to PageRank. However, we calculate ranks on an integrated space of the web and a database with additional consideration of semantic links, as opposed to the PageRank's approach. Recently, [3, 4] consider relationships between database entries as a kind of links and apply a link analysis to a database. We extend this idea to perform a link analysis based on the integrated graph of the web and a user database.

Web and its geographical locality There exist many approaches for extracting pages related to a specific geographical area from the web. [9] proposes a notion of a *localness degree* to discover local information from the web. An *augmented web space* is presented in [10]. This augmented web space consists of web pages, hyperlinks and semantic links that represent geographical relationships between web pages. In contrast to these researches, we consider geographical locality of a web page by employing virtual geographical links to connect the web and a database and applying them a link analysis. The approach proposed in [11] is to determine *geographical scopes* of web resources. Based on the textual contents of a web resource, as well as the distribution of related hyperlinks, the scope of the resource is computed. In [12], a categorization method of queries to a search engine is proposed. Its feature is that the categorization is performed based on geographical locality. Some of these approaches may be helpful to extend our method.

Topic-focused crawling Although it is not our emphasis to crawl web pages related to user database entries, our method is related to the web crawling technology. Recently, *topic-focused crawling* [13–16] is becoming a key technology for efficiently collecting web pages. [13] develops a framework to evaluate topic-focused crawling algorithms. The *PageRank crawler* [14] prefers accessing a web page with a high PageRank value. The *focused crawler* in [15] is based on a hypertext classifier. Its basic idea is to classify crawled pages with categories in a topic taxonomy. The *context focused crawler* in [16] guides its crawl using Bayesian classifiers trained to estimate the link distance between a crawled page and the relevant target pages. Our LocalRank values can be used to navigate a crawler to find highly relevant pages to a database, like a PageRank crawler. In this sense, our approach can be applied to a topic-focused crawling, where the "topic" is determined by database entries.

3 Proposed Method

In this section, we describe our method using the example table of restaurants in Tsukuba (Fig. 1). Subsection 3.1 presents the concept of an extended database. Subsection 3.2 shows a model for integrating the web and a database. Subsection 3.3 introduces how a graph structure is constructed and how LocalRank is calculated.

3.1 Extended Database

For a database of Fig. 1, we first consider its *extended database* shown in Fig. 2. This extended database, whose left and right parts denote the database and the web respectively, shows their relationships in the style of the entity-relationship diagram. The `restaurant` entity represents the set of the entries of Fig. 1. The `page` entity stands for the total set of all the web pages. The web pages' URLs, addresses, phone numbers and zip codes appearing on them are considered as the attributes of the `page` entity. The lines with two arrows show that an attribute may have zero to many values, because it is possible that zero to many addresses appear on a web page. The `has-HP` relationship connects the `restaurant` and `page` entities. The notations "(1, 1)" and "(0, 1)" mean that for a certain `restaurant` entry there must exist a corresponding homepage, while not all web pages must have a corresponding `restaurant` entry. The `refers` relationship represents hyperlink references between web pages.

Notice that for other databases, extended databases can similarly be constructed using additional attributes. Therefore attributes are not restricted to addresses, phone numbers, etc. They may be the ones that can be extracted from web pages with available tools according to user requirements.

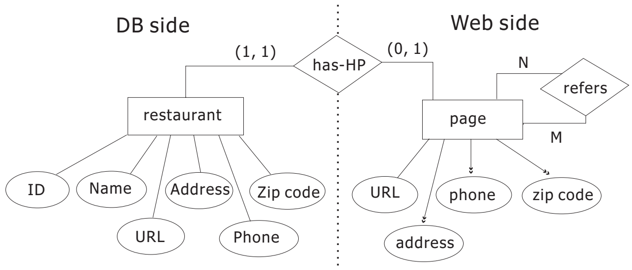

Fig. 2. Extended database

3.2 Authority Transfer Graph

We describe next an *authority transfer graph* (Fig. 3), based on the idea of [3]. This graph is constructed based on the extended database explained in the previous subsection and reflects some relationships between its elements. For simplicity the attributes are omitted.

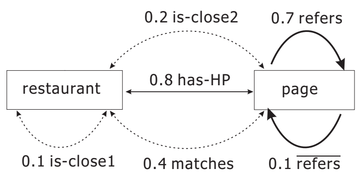

Fig. 3. An example of authority transfer graph

We call the solid lines and dotted lines in this graph *links*, since we extend the concept of hyperlinks between web pages to the integrated space of the web and a database. We also call the values between 0 and 1 assigned to links their *weights*. Actually, how to assign proper weights to links is a difficult task. It may be a trial process performed until the quality of results becomes satisfactory. In this paper we do not give more emphasis to the discussion on this problem. The solid lines correspond to the relationships `has-HP` and `refers` in the extended database (Fig. 2). The `refers` link means a reference from a web page to another page and the weight "0.7" means that a page's score is multiplied by 0.7 and then transferred to its linked page. The $\overline{\texttt{refers}}$ link represents a link in the opposite direction. In this example, the weight "0.1" means that the score of a link target page is multiplied by 0.1 and then transferred to the page linking to it. A kind of links, whose weights in both directions are equal, is represented by $\longleftrightarrow$. For example, we use the `has-HP` link to represent a correspondence relationship between a database entry and its homepage, and assume the weights from any side to another side are equal (i.e., they influence each other equally). In this case we use $\longleftrightarrow$ for `has-HP`.

The dotted lines show semantic relationships. The `is-close1` link is generated between two `restaurant` entries when their addresses are close. The `is-close2` link connects `restaurant` and `page` entities, and is generated when their geographical locations are close. The `matches` link is generated when it can be determined that the information a web page describes is about a `restaurant` entry. In general, there often exists flexibility in semantic links. Therefore, we assume that a user-defined predicate is given to define a semantic link. For example, `is-close1` is generated when the distance between two restaurant addresses is below a certain threshold. We will later describe its sample implementation for the experiment. Although the `is-close1` link in this example is based on a binary decision (i.e., an `is-close1` link is generated or not), we may be able to assign more sophisticated weights to this kind of links, say, depending on their distance values. Such issues are our future challenges.

3.3 Link Analysis

In our method, we first construct a *data graph* (Fig. 4) in which the links defined in the authority transfer graph (Fig. 3) are realized. The data graph consists of database entries, their related pages, hyperlinks, and some semantic links. Then we perform a link analysis based on this data graph. The result of this link analysis, called *LocalRank*, combines the web pages' popularity on the local web space with their relevance degrees to the database entries. The LocalRank score is influenced by geographical locality because we integrate the web and a user database using semantic links including virtual geographical links (`is-close1` and `is-close2`).

The following steps are what we have actually done in the experiment. This process may be able to be extended depending on the environment and the requirement when we apply our approach to a different context.

Constructing a data graph First we describe how to generate the nodes of the data graph.

1. For each entry of the `restaurant` table, generate a corresponding node.
2. Download the homepage of each restaurant using the "URL" attribute of the table and let the set of pages be S_1.
3. Get the web pages that are reachable from the homepages in the same web site[1] and let the set be S_2.
4. For each homepage, retrieve its backlink pages using a search engine (Google is used in our experiment). These web pages compose a set S_3.
5. Perform keyword searches using the attribute values of each row of the `restaurant` table. In the experiment, we retrieve web pages using the "Name" and "Phone" attributes to create a query condition. The result set of web pages is called S_4.
6. For all the web pages in S_1, S_2, S_3 and S_4 acquired in Step 2, 3, 4 and 5, generate their corresponding nodes.
7. Extract URLs from the web pages in S_1, S_2, S_3 and S_4 and generate nodes for those whose corresponding pages have not been downloaded.

As shown above, the nodes of the data graph consist of the entries of the `restaurant` table and their related web pages including the downloaded and undownloaded ones.

Next the links between these nodes are generated as follows.

1. For the web pages in S_1, connect them to the corresponding `restaurant` entries with bidirectional `has-HP` links.

[1] The decision is simply based on judging whether the URL string of a web page has a same prefix with that of a homepage.

2. Based on the results of the URL extraction in Step 7 of the node generation, create `refers` and $\overline{\texttt{refers}}$ links between web pages (including the downloaded and undownloaded ones).
3. For the web pages in S_4 (obtained from a search engine using the "Name" and "Phone" attributes), connect them to the corresponding `restaurant` entries with bidirectional `matches` links.
4. Generate bidirectional `is-close1` links between database entries when the user-defined `is-close1` predicate is true. In the experiment, we implement the predicate by mapping the "Address" attributes of the database entries to coordinates and judging whether their distances, calculated using the coordinates, are below a given threshold or not.
5. Similarly, generate bidirectional `is-close2` links between the web pages in $S_1 \cup S_2 \cup S_3 \cup S_4$ and their corresponding database entries if the `is-close2` predicate is true.

We assign weights to the links based on the settings of the authority transfer graph. Notice that when two or more edges belonging to a same kind of links go out from a same node, the weight is divided by the number of outgoing edges. Let us see the example of Fig. 4. Since there are two `matches` links going out from the COCCOLINO entry, the weight "0.4" of the `matches` link in the authority transfer graph (Fig. 3) is divided by two and the weights of both outgoing edges become "0.2". Meanwhile the weight of the edge from a corresponding web page to the COCCOLINO entry remains "0.4". This weight setting approach is similar to that of PageRank [8].

As mentioned above, a graph structure called a *data graph* (Fig. 4) is constructed. In Fig. 4, the left oval nodes represent the entries of the `restaurant` table, while the middle square nodes and right circle nodes denote the downloaded web pages and the undownloaded URLs, respectively. The labels of the `refers` and $\overline{\texttt{refers}}$ links are omitted.

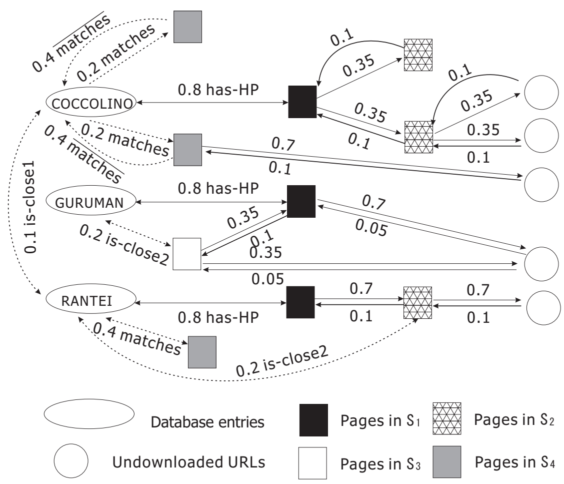

Fig. 4. An example of data graph

Score calculation We calculate next the score called *LocalRank* for each node[1] in the data graph constructed in the previous section. It can be calculated by the following equation.

$$\boldsymbol{r} = d\mathbf{A}'\boldsymbol{r} + \frac{(1-d)}{n}\boldsymbol{e} \quad (1)$$

where n is the number of the nodes in the data graph, d is a damping factor given by a user, $\mathbf{A}$ denotes an $n \times n$ matrix, the value of whose element (i, j) sums the weights of all edges from i to j, $\mathbf{A}'$ is the transposed matrix of $\mathbf{A}$, $\boldsymbol{r}$ denotes a column vector with n dimensions, whose element is the LocalRank value of each node, and $\boldsymbol{e}$ denotes an n-dimension column vector with all elements set to 1. All elements of $\boldsymbol{r}$ are initialized to 1 and the calculation terminates when $\boldsymbol{r}$ converges.

Notice that although the calculation equation is similar to PageRank, the sense of our LocalRank is different from PageRank. Since we calculate the scores based on the data graph which integrates the web and a user database, instead of the conventional web space, our LocalRank reflects how a web page is related to the database entries which the user provides, in addition to its popularity on the web space. Also notice that instead of the global web space, a local web space is used to perform a link analysis. Hence, our method ranks web pages considering their relevance degree to user's interest. Moreover, let us remind that we have used the geographical `is-close1` and `is-close2` links to connect the web and the database and applied them to the link analysis. Therefore, geographical locality is also infused into the rankings of web pages.

4 Experiments

4.1 Database Construction and Collection of Related Pages

We select fifty-four homepages of restaurants in Tsukuba city from gourmet navigation pages (e.g., [17]) of some portal web sites about Tsukuba city. We also manually extract from these pages restaurants' names, URLs, and addresses and construct a database as shown in Fig. 1 using them as database attributes.

The pages related to the database are collected by the approach described in the previous section. They are 54 homepages (S_1), 1,409 web pages (S_2) reachable from the homepages, 166 backlink pages (S_3) of the homepages, and 285 web pages (S_4) acquired from Google using the "Name" and "Phone" attributes (e.g., "COCCOLINO" and "029-864-4555") as keywords. After the elimination of duplicate ones, 1,812 web pages can be obtained.

4.2 Information Extraction and Data Graph Construction

Then the useful information is extracted from the downloaded web pages by the following steps.

1. Link extraction: Extract all URLs except for those referring to media files or CGI files.
2. Zip code extraction: Extract zip codes, like "305-8573" (the hyphen can be omitted), from the web pages containing only one zip code. We do not use pages on which two or more zip codes appear because of their vagueness.

[1] Although we calculate scores for all nodes, we are much more interested in the ranks of web pages than those of database entries because we aim to effectively rank pages so as to find relevant ones according to user's interest.

3. Address extraction: Extract full address descriptions, like “1-1-1 Tennodai, Tsukuba, Ibaraki” from a web page if it contains only one address. In this step, partial address descriptions (e.g., “Tsukuba, Ibaraki”) are not used due to their vagueness.
4. Coordinate calculation: Calculate their corresponding latitudes and longitudes from the addresses acquired in Step 3 using Yahoo!JAPAN MAPS [18]. Given an address or zip code, this service returns a pair of latitude and longitude. For the “Address” attribute value of each row in Fig. 1, their coordinates are also calculated.

Then nodes and links of a data graph are generated. This data graph has 13,006 nodes consisting of 54 database entries and 12,952 web pages including 1,812 downloaded and 11,140 undownloaded ones. Additionally, 54 has-HP, 8,663 refers, 8,663 $\overline{\texttt{refers}}$, 285 matches, 453 is-close1, and 1,407 is-close2 are generated, respectively. We connect two database entries with the is-close1 links when 1) the distance of their addresses is smaller than 2km, or 2) their zip codes are same. We calculate an accurate distance based on the system [19] serviced by Geographical Survey Institute, Japan. The is-close2 links are similarly generated between a web page and a database entry considering addresses or zip codes appearing on a web page.

4.3 Link Analysis Results

For the LocalRank calculation, we use a free scientific software “Scilab” [20], which has a sparse matrix handling facility, and let d in Equation (1) be 0.9. The information of the 20 top-ranked downloaded pages is shown in Fig. 5. We evaluate two features of a page by manually examining its content. The first one is whether a page describes information on restaurants or not. The second one is whether it presents information on the area of Tsukuba city or not. We mark a page with “Y” if it is highly relevant and “N” if completely unrelated. The symbol “B” denotes a border page (i.e., it is somewhat relevant). Pages 1-5, 8-15, 18 in Fig. 5 are the ones related to the “RANTEI” restaurant which is an entry of Fig. 1. Since this restaurant is linked from the major local portal sites and restaurant navigation sites, its related web pages are highly ranked. Page 20 is the homepage of the “WHITE-GYOUZA” restaurant that is also an entry of our sample database. Pages 16 and 17 are two link collections of restaurants in Tsukuba and its adjacent cities. Among the 20 pages, 17 pages describe the information on restaurants and/or the area of Tsukuba. Only three pages have neither of these two features. The web pages ranked after the top-20 ones are omitted here, but they have similar tendency.

Fig. 6 shows the 20 top-ranked undownloaded URLs. These web pages may be considered to be good candidates to begin the crawling to collect related web pages to the database. Pages 1, 8-11 are portal sites for Tsukuba which introduce general information including restaurant information on this city. Pages 2-4 are from a popular restaurant navigation site in Japan. In particular, page 3 focuses on the area of Ibaraki prefecture that includes Tsukuba city. In this sense, its mark of the “Tsukuba” column is denoted as “B”. Pages 12-20 are the maps of some restaurants. Among the 20 pages, 17 pages either are related to restaurant information, or introduce the information on Tsukuba. Only three remaining pages (Pages 5, 6, and 7) are completely undesirable. The web pages corresponding to the URLs following this top-20 list also contain restaurant pages and local web pages, but their quality gradually goes down as ranks become lower. Based on this experiment, we can say that our LocalRank values are semantically meaningful since they reflect our intent to obtain highly related pages to the given local restaurant database. The rankings can serve as a reference when a user wants to select related pages to access from the undownloaded URLs.

Ranking	LocalRank	URL	Restaurant	Tsukuba
1	0.260874	e-tsukuba.jp/rantei/index.htm	Y	Y
2	0.021487	r.gnavi.co.jp/a275100/	Y	Y
3	0.018932	e-tsukuba.jp/rantei/link.htm	Y	Y
4	0.014155	www.geocities.jp/papy0164/syokuji/you.html	Y	B
5	0.012105	r.gnavi.co.jp/a275100/map1.htm	Y	Y
6	0.012086	www.collaborate-ibaraki.jp/dirsearch/kigyou/namediv/ni.asp	N	N
7	0.012056	www.joyoliving.co.jp/kurashi/data/thisweek.php?category=recruit	N	N
8	0.011492	www.joyo-net.com/mise/mise040408.html	Y	Y
9	0.011492	www.capital-group.co.jp/rantei.htm	Y	Y
10	0.011492	www.piazza.ne.jp/piazza/gourmet/index.asp?mode=detail&id=116	Y	Y
11	0.007308	e-tsukuba.jp/rantei/recruit.htm	Y	Y
12	0.007285	e-tsukuba.jp/rantei/company.htm	Y	Y
13	0.007185	e-tsukuba.jp/rantei/osusume.htm	Y	Y
14	0.007185	e-tsukuba.jp/rantei/email.htm	Y	Y
15	0.007185	e-tsukuba.jp/rantei/traffic.htm	Y	Y
16	0.002886	www.iki-iki.net/v7/townpage/a-you.htm	Y	B
17	0.002830	tarea.hp.infoseek.co.jp/lminami.html	Y	B
18	0.002737	www006.upp.so-net.ne.jp/puni/ibaraki-r-w1.htm	B	B
19	0.002670	www.h3.dion.ne.jp/ b-gakuji/syaon.html	N	N
20	0.002266	www.white-gyouza.co.jp/detail/detail18.htm	Y	Y

Fig. 5. 20 top-ranked downloaded pages

5 Conclusion and Future Work

In this paper, we present a method called LocalRank to rank web pages by integrating the web and a local user database. The experimental results show that our framework is helpful. Notice that our method is flexible and may be applied to the collection and ranking of web pages for other user data, not restricted to the local restaurant information as mentioned in this paper, just by changing the type of semantic links and properly adjusting their weights.

We believe that there are still a number of interesting work that need to be carried out. First, we need to extend our experiment to a larger user database, and to different geographical areas. Moreover, it would be desirable to further investigate the influence of changing the weights of links. In addition, it is also necessary to pick up useful records from the collected web pages and complement the user database with them. We may be able to consider to use the LocalRank value as a reference to judge whether a page is useful for a database expansion.

Acknowledgement

This research is partly supported by the Grant-in-Aid for Scientific Research (16500048) from Japan Society for the Promotion of Science (JSPS), Japan and the Grant-in-Aid for Scientific Research on Priority Areas (16016205) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. In addition, this work is supported by the grants from the Asahi Glass Foundation and the Inamori Foundation.

References

1. Regional in the Yahoo! Directory. `http://dir.yahoo.com/Regional/`
2. goo TOWNPAGE. `http://townpage.goo.ne.jp/`
3. A. Balmin, V. Hristidis, and Y. Papakonstantinou, ObjectRank: Authority-Based Keyword Search in Databases. *Proc. VLDB 2004*, pp. 564-575, 2004.

Ranking	LocalRank	URL	Restaurant	Tsukuba
1	0.019240	www.i-tsukuba.com/index.shtml	B	Y
2	0.002976	www.gnavi.co.jp	Y	N
3	0.002976	www.gnavi.co.jp/kanto/	Y	B
4	0.002924	my.gnavi.co.jp/Rating/regist.php?shopid=a275100&shopurl...	Y	N
5	0.002304	www.nilim.go.jp	N	N
6	0.002304	www.icube-t.co.jp	N	Y
7	0.002304	www.google.co.jp/custom	N	N
8	0.001833	www.tsukuba.ad.jp	B	Y
9	0.001817	www.i-tsukuba.com	B	Y
10	0.001815	www.ibarakiken.net	B	B
11	0.001813	www.e-tsukuba.jp	B	Y
12	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
13	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
14	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
15	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
16	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
17	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&t=s	B	B
18	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
19	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B
20	0.000468	rm.gnavi.co.jp/Map/mc_view.php?dr=a275100&c=36...	B	B

Fig. 6. 20 top-ranked undownloaded URLs

4. F. Geerts, H. Mannila, and E. Terzi, Relational Link-Based Ranking. *Proc. VLDB 2004*, pp. 552-563, 2004.
5. P. Baldi, P. Frasconi, and P. Smyth, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, Wiley, 2003.
6. S. Chakrabarti, *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann, 2002.
7. J.M. Kleinberg, Authoritative Sources in a Hyperlinked Environment. *JACM*, Vol. 46, No. 5, pp. 604-632, 1999.
8. S. Brin and L. Page, The Anatomy of a Large-scale Hypertexual Web Search Engine. *Computer Networks*, Vol. 30, No. 1-7, pp. 107-117, 1998.
9. C. Matsumoto, Q. Ma, and K. Tanaka, Web Information Retrieval Based on the Localness Degree, *Proc. DEXA 2002*, LNCS 2453, pp. 172-181, 2002.
10. K. Hiramatsu and T. Ishida, An Augmented Web Space for Digital Cities. *Proc. SAINT 2001*, pp. 105-112, 2001.
11. J. Ding, L. Gravano and N. Shivakumar, Computing Geographical Scopes of Web Resources. *Proc. VLDB 2000*, pp. 545-556, 2000.
12. L. Gravano, V. Hatzivassiloglou and R. Lichtenstein, Categorizing Web Queries According to Geographical Locality. *Proc. CIKM 2003*, pp. 325-333, 2003.
13. F. Menczer, G. Pant, and P. Srinivasan, Topical Web Crawlers: Evaluating Adaptive Algorithms. *ACM Trans. on Internet Technology*, Vol. 4, No. 4, pp. 378-419, 2004.
14. J. Cho, H. Garcia-Molina, and L. Page, Efficient Crawling through URL Ordering. *Computer Networks*, Vol. 30, No. 1-7, pp. 161-172, 1998.
15. S. Chakrabarti, M. van den Berg, and B. Dom, Focused Crawling: A New Approach to Topic-specific Web Resource Discovery. *Computer Networks*, Vol. 31, No. 11-16, pp. 1623-1640, 1999.
16. M. Diligenti, F. M. Coetzee, S. Lawrence, C.L. Giles, and M. Gori, Focused Crawling Using Context Graphs. *Proc. VLDB 2000*, pp. 527-534, 2000.
17. Gourmet navigation in Tsukuba city. `http://www.i-tsukuba.com/yellow/gourmet/index.htm`
18. Yahoo! JAPAN MAPS. `http://map.yahoo.co.jp/`
19. Measurement Calculation. `http://vldb.gsi.go.jp/sokuchi/surveycalc/bl2stf.html`
20. Scilab Home Page. `http://scilabsoft.inria.fr/`